

Lecture Notes  
Lectures 14, 15, 16, & 17  
*CS389: Machine Learning, Spring 2023*  
*University of Massachusetts Amherst*

## 1 Reinforcement Learning (lecture 14)

### 1.1 Abstract Overview (What is the goal? and when to use it?)

Reinforcement learning concern itself with the study of learning and intelligence. Think of how you would train your pet dog (or dragon – great series) to perform a skill. It is likely that the dog learns through positive or negative reinforcement training from rewards such as treats, praise, toys, etc. *We will formalize this intuition mathematically in the following sections.* But before we start, it is important to distinguish between RL and supervised learning. A common misconception is that RL is an alternative to supervised learning. This is possible but should not be done. If you have labels for your data, do not discard them and convert the feedback from instructive feedback (telling the agent what label it should have given) to evaluative feedback (telling the agent if it was right or wrong). It is likely that RL methods will perform far worse so you should only resort to RL when you have a sequential decision problem or a problem where only evaluative feedback is available [4].

### 1.2 Describing the Agent and Environment Mathematically

On a high level, we want to find an optimal policy that maximizes the expected total amount of reward the agent will obtain. In order to apply this intuition of learning to our problems, we need the ability to precisely describe both the agent, environment and their interactions.

To reason about learning, we will need to mat

An example of an *agent* include a child, dog, self-driving car, etc.

### 1.2.1 Agent's Goal

Objective Function

Optimal Policy

Reward Discounting

## 2 Actor-Critic (lecture 15)

2.1 Dopamine and TD Error

2.2 Value Estimation and TD Error

2.3 Policy Gradients

## 3 Q-Learning (lecture 16)

## 4 Q-Estimation (lecture 17)

## References

- [1] Cooper.
- [2] Fei-Fei Li, Jiajun Wu, and Ruohan Gao, Stanford CS231n, Spring 2022. <https://web.archive.org/web/20230109135558/https://cs231n.github.io/>
- [3] Graphs of tanh, sigmoid, ReLU. <https://studymachinelearning.com/activation-functions-in-neural-network/>
- [4] Bruno Castro da Silva and Philip Thomas, COMPSCI 687 Reinforcement Learning Lecture Notes (Fall 2022) [https://people.cs.umass.edu/~bsilva/courses/CMPSCI\\_687/Fall2022/Lecture\\_Notes\\_v1.0\\_687\\_F22.pdf](https://people.cs.umass.edu/~bsilva/courses/CMPSCI_687/Fall2022/Lecture_Notes_v1.0_687_F22.pdf)